PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-051890
(43)Date of publication of application : 23.02.2001

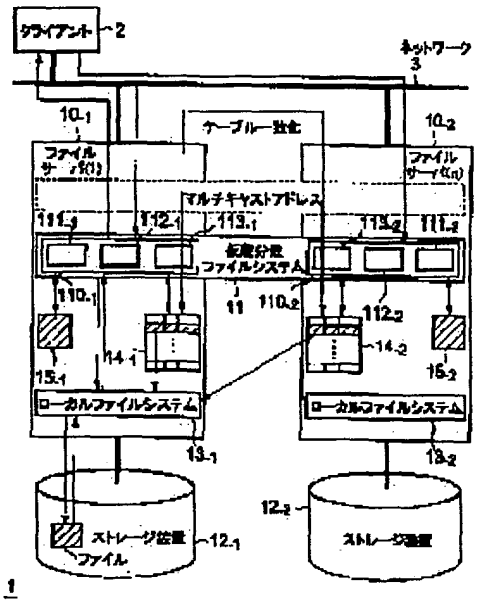
(51)Int.Cl. G06F 12/00
G06F 13/00

(21)Application number : 11-226494
(22)Date of filing : 10.08.1999

(71)Applicant : TOSHIBA CORP
(72)Inventor : UCHIBORI IKUO
TAKAKUWA MASAYUKI

(54) VIRTUAL DECENTRALIZED FILE SERVER SYSTEM

(57)Abstract:
PROBLEM TO BE SOLVED: To make a client not pay attention to the number of file servers decentralized in a network and the *connection states of storage devices.*
SOLUTION: This virtual decentralized file server system 1 is equipped with servers 10-1 and 10-2 decentralized in the network 3 and a virtual decentralized file system 11 is decentralized and mounted on each of the servers. Modules 110-1 and 110-2 on the servers 10-1 and 10-2 which constitute this system 11 when receiving a file operation request multicast from a client 2 judge whether or not their servers are optimum servers capable of handling the request according to server information holding parts 15-1 and 15-2 holding mapping tables 14-1 and 14-2 between the virtual decentralized file system 11 and all local file systems 13-1 and 13-2 or server information on all the servers, and makes a local file system of a corresponding server perform requested file operation according to the judgement result.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

* NOTICES *

1. This document has been translated by computer.So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3.In the drawings, any words are not translated.

CLAIMS

[Claim(s)]
[Claim 1] It has the following. the aforementioned virtual distributed file system It consists of management modules formed in each aforementioned file server, respectively. each aforementioned management module By receiving in common the file manipulation demand multicasted from the client, and referring to the aforementioned mapping table of a self-server according to the demand concerned It judges whether it is the optimal server to which a self-server can process the demand concerned. The virtual distribution file server system characterized by being constituted so that the aforementioned corresponding local file system of a server may be made to perform demanded file manipulation, only when it is judged that it is the optimal server. The virtual distributed file system of not being dependent on the actual storage composition which is the virtual distribution file server system equipped with two or more file servers distributed on the network which can be multicasted, is distributed and mounted in each aforementioned file server, and manages the file of all file servers in integration. The local file system which is mounted independently in each aforementioned file server, respectively, and manages storage composition peculiar to each server. The mapping table which holds the information on mapping between the virtual distribution file server system concerned and the aforementioned local file system which actually manages the file about each file which is prepared in each aforementioned file server, respectively, and is managed in integration by the aforementioned virtual distributed file system.
[Claim 2] The virtual distribution file server system equipped with two or more file servers distributed on the network which can be multicasted characterized by providing the following. The virtual distributed file system of not being dependent on the actual storage composition which is distributed and mounted in each aforementioned file server, and manages the file of all file servers in integration. The local file system which is mounted independently in each aforementioned file server, respectively, and manages storage composition peculiar to each server. The mapping table which holds the information on mapping between the virtual distribution file server system concerned and the aforementioned local file system which actually manages the file about each file which is prepared in each aforementioned file server, respectively, and is managed in integration by the aforementioned virtual distributed file system. Either [at least] the information which is prepared in each aforementioned file server, respectively, and shows the availability of the storage equipment of the server about all the aforementioned file servers, or the information which shows the load situation of the server.
[Claim 3] The aforementioned management module is the virtual distribution file server system according to claim 1 or 2 it carries out [whether when the aforementioned file manipulation demand is a file read-out demand or a file write request, with reference to the aforementioned mapping table of a self-server, the corresponding file is under management of the aforementioned local file system of a self-server, and] whether it is the optimal server to which a self-server can process the aforementioned demand, and that it judges as the feature.
[Claim 4] It is the virtual distribution file-server system according to claim 2 characterized by to judge whether the aforementioned management module is comparing the availability of the storage equipment of the server, or the load situation of the server about each of all the aforementioned servers with reference to the aforementioned server information maintenance means of a self-server, and is the optimal server to which a self-server can process the aforementioned demand when the aforementioned file manipulation demand is a new creation demand of a file.
[Claim 5] The aforementioned management module is a virtual distribution file server system according to claim 1 characterized by exchanging the information on the aforementioned mapping table of a self-server, and the information on the aforementioned mapping table of other servers by communication between servers in order to carry out identification of the content of the aforementioned mapping table of all the aforementioned file servers.
[Claim 6] In order that the aforementioned management module may carry out identification of the content of the aforementioned mapping table of all the aforementioned file servers While exchanging the information on the aforementioned mapping table of a self-server, and the information on the aforementioned mapping table of other servers by communication between servers, in order to carry out identification of the content of the aforementioned server information maintenance means of all the aforementioned file servers The virtual distribution file server system according to claim 2 characterized by exchanging the information on the aforementioned server information maintenance means of a self-server, and the information on the aforementioned server information maintenance means of other servers by communication between servers.
[Claim 7] It is prepared in each aforementioned file server, respectively, and a load status information maintenance means classified by file to hold the information which shows the load situation according to each file under management of the server is provided further. The aforementioned management module detects the file of the load which exceeded the 1st threshold from the information currently held at the aforementioned load status information maintenance means classified by file of a self-server. Communication between servers performs the replication of the file concerned to other arbitrary file servers. The virtual distribution file server system according to claim 1 or 2 characterized by leaving the processing to the demand concerned to a replication side when there is a read-out demand of the file concerned multicasted from the client.

[Translation done.]

http://www4.ipdl.jpo.go.jp/cgi-bin/tran_web_cgi_ejje?u=http%3A%2F%2Fwww6.ipdl.jp... 2003/07/22

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[The technical field to which invention belongs] this invention carries out cooperation operation of two or more file servers which started the file server system in a computer network system, especially were connected on the network, and relates to the virtual distribution file server system operated as a single server from a client.
[0002]
[Description of the Prior Art] Generally in today's computer network system, sharing a file between different computers connected to the network is performed. Under such environment, large-scale storage is connected to a specific computer, it applies as a file server or, recently, the system configuration of connecting the file server special-purpose machinery called NAS (Network Attached Storage) is taken in many cases.
[0003] In the environment (file server system) which uses a file server, if expandability is in a server side physically and efficiently when the storage capacity of a server runs short, it can be coped with by newly extending a disk unit etc. (storage equipment). At this time, it becomes the gestalt of using it, mounting new volume, from a client. Moreover, the server itself will be extended if the expandability of a server has reached the limitation. At this time, it becomes the gestalt of using it, mounting new volume after he is conscious of the extended server from a client.
[0004]
[Problem(s) to be Solved by the Invention] When performing file sharing in the above-mentioned computer network system using a file server, it is common that the volume composition by the side of a file server is in sight as it is from a client side. For example, when a disk unit is extended by the server side, when new volume has been recognized, you have to mount a client side. Or when the server itself is extended, the complicated work of determination or system construction, management, etc. generates the employment policy of the extended server, and the new server has been recognized also by the client side, you have to mount new volume.
[0005] Thus, in the file-sharing system (file server system) using the conventional file server, when extension of a disk unit (storage equipment) or extension of a server was required, the problem that cost great for new setup and management occurred was in all of a client side the server side. Furthermore, there was a use gestalt of storage, when only capacity wanted to extend a specific file system as it is, and it also had the case which is not solved only by extending storage equipment and a server.
[0006] this invention was made in consideration of the above-mentioned situation, and the purpose can treat from a client two or more file servers distributed on the network as a single server, and is to offer the virtual distribution file server system which does not make a client conscious of the connection state of the number of a server, or storage equipment.
[0007]
[Means for Solving the Problem] The virtual distributed file system of not being dependent on the actual storage composition which this invention is distributed and mounted in two or more file servers connected to the network which can be multicasted, and manages the file of all file

servers in integration, The local file system which is mounted independently in each file server, respectively, and manages storage composition peculiar to each server, It is prepared in each aforementioned file server, respectively. about each above-mentioned file Information on mapping between a virtual distribution file server system and the local file system which actually manages the file (for example, it is managed by the virtual distribution file server system, and with the imagination path which appears from a client) While having a mapping table holding the information which matched the physical whereabouts which is managed with a local file system and is not visible from a client It is the management module in which the above-mentioned virtual distributed file system was prepared by each file server, respectively. By receiving in common the file manipulation demand multicasted from the client, and referring to the mapping table of a self-server according to the demand concerned Only when a self-server judges whether it is the optimal server which can process the demand concerned and judges that it is the optimal server, it is characterized by constituting with the management module to which the demanded file manipulation is made to perform with the corresponding local file system of a server.
[0008] The information which shows the availability of the storage equipment of the server about all file servers on each file server here, A server information maintenance means to hold the server information containing at least one side of the information which shows the load situation of the server is established further. and by each above-mentioned management module When the file manipulation demand multicasted from the client is received, it is good also as composition which judges whether it is the optimal server to which a self-server can process the demand concerned by referring to the mapping table of a self-server, or a server information maintenance means according to the demand concerned.
[0009] In such composition, the file manipulation demand multicasted without being conscious of a specific file server from a client is received in common by the management module on each file server which constitutes a virtual distribution file server system, and the mapping table of a server (self-server) which corresponds according to the demand, or a server information maintenance means is referred to. And only the only server (upper management module) which it was judged whether it is the optimal server to which a self-server can process the above-mentioned demand, and was judged to be the optimal server makes the local file system of a self-server perform demanded file manipulation as a result of this reference.
[0010] Thus, from the client of a requiring agency, he can treat two or more file servers distributed on the network as a single server, and does not need to be conscious of the connection state of the number of a server, or storage equipment.
[0011] It is good to apply either the following 1st or the 4th algorithm (the judgment technique) as an algorithm for judging whether a self-server is optimal server by the above-mentioned management module here.
[0012] The 1st algorithm is technique judged by whether the file which is applied when a file manipulation demand is a file read-out demand or a file write request, and corresponds based on the information on the mapping table of a self-server is under management of the local file system of a self-server.
[0013] The 2nd algorithm is what is applied when a file manipulation demand is a new creation demand of a file. It is based on the information on the server information maintenance means of a self-server. about each of all servers It is the technique (for example, when the availability of a self-server is the largest, the load of a self-server judges it as the above-mentioned optimal server to a low case most) judged by comparing the availability (empty storage capacity) of the storage equipment of the server, or the load situation of the server.
[0014] The 3rd algorithm is also the technique (for example, when the size of a continuation field securable on the storage equipment of a self-server is the largest, it judges as the above-mentioned optimal server) which judges by being applied when a file-manipulation demand is a new creation demand of a file, asking for a continuation field securable on the storage equipment which corresponds about each of all servers based on the information on the mapping table of a self-server, and comparing the size of the continuation field.
[0015] The 4th algorithm is also the technique judged by being applied when a file manipulation

demand is a new creation demand of a file, calculating at least two, the availability of the storage equipment of the server, the load of the server, and a continuation field securable on the storage equipment concerned, about each of all servers, and comparing the at least two information searched for as compound condition.
[0016] Even if each server receives in common the file manipulation demand multicasted without being conscious of a specific file server from a client by applying any one of the 1st of a more than, or the 4th algorithm, it can judge autonomously whether it is the optimal server for performing the demanded file manipulation for the server itself, without communicating mutually each time.
[0017] It is good to give the function (communication module) to exchange the information on the mapping table of a self-server and the information on the mapping table of other servers by communication between servers to each above-mentioned management module here, in order to carry out identification of the content of the mapping table of all file servers. Moreover, it is good to give the function to exchange the information on the server information maintenance means of a self-server and the information on the server information maintenance means of other servers by communication between servers, further to each management module (inner communication module) with the composition which was equipped with the server information maintenance means on each server in addition to the mapping table, in order to carry out identification of the content of the server information maintenance means of all file servers.
[0018] Moreover, for the identification of a mapping table, when the file organization actually managed with the local file system of a self-server is changed, it is efficient to transmit the changed information (mapping information) to all other servers by communication between servers. Similarly, for the identification of the content of a server information maintenance means, it is efficient to update the server information on a self-server periodically, and to transmit the updated server information to all other servers by communication between servers each time.
[0019] Moreover, this invention is characterized also by performing the 1st of the following [the management module of the server], or 4th processing, when a server is dynamically extended by the above-mentioned virtual distribution file server system. In the 1st processing, a lock setup which forbids renewal of a mapping table and a server information maintenance means to all other servers by communication between servers is performed. first, in the following processing of the 2nd The content of a mapping table server information maintenance means is copied to a self-server from other arbitrary servers by communication between servers. in the following processing of the 3rd The server information on a self-server is added to the server information maintenance means of a self-server. in the following processing of the 4th - BA information on a self-server is made to reflect in the server information maintenance means of all other servers by communication between servers, identification of the server information maintenance means of all servers is attained, and the above-mentioned lock setup is canceled after an appropriate time.
[0020] The number of a server is dynamically extensible with a series of operation at the time of such server extension. And a client can use the extended server, without being conscious of extension of the number of a server.
[0021] Moreover, this invention adds a load status information maintenance means classified by file to hold the information which shows the load situation according to each file under management of the server to each file server, and sets it to the management module of each server. The file of the load which exceeded the 1st threshold from the information currently held at the load status information maintenance means classified by file of a self-server is detected. Communication between servers performs the replication of the file concerned to other arbitrary file servers. When there is a read-out demand of the file concerned multicasted from the client, it is characterized also by leaving the processing to the demand concerned to a replication side.
[0022] In such composition, an autonomous load distribution becomes possible. Here, an autonomous load distribution can carry out to a large area more effectively by communication between servers performing the replication of the file concerned to other arbitrary file servers, when the file which carried out [above-mentioned] detection is a file by which the replication

was carried out, and leaving the processing to the demand concerned to a new replication side, when there is a read-out demand of the file concerned multicasted from the client. Moreover, when the load of a file set as the object of the replication to other servers consists of the 1st threshold of the above below the threshold of a low 2nd Elimination of the file which corresponds by communication between servers to the other servers concerned from the management module on the server which performed the replication is required. When there is a read-out demand of the file concerned multicasted from the client, and self performs processing to the demand concerned, a dynamic load distribution becomes possible.
[0023] Now, it is possible the communication between servers for the above-mentioned identification (communication of mapping information or server information) and to use the above-mentioned network for communication between servers for the above-mentioned replication further. However, it is good to consider as the composition which forms the channel (private channel) of the exclusive use which interconnects each file server independently of the above-mentioned network, and performs communication between servers using the channel concerned. In this case, it can prevent that the throughput of a network gets worse for communication between servers.
[0024] Moreover, this invention is characterized also by performing the above-mentioned communication between servers by each above-mentioned management module through the interface concerned while it is further equipped with the interface in which the multi-host who interconnects the storage equipment of each file server and the server concerned is possible.
[0025] In such composition, since each above-mentioned storage equipment is shared between each server with the above-mentioned interface and communication between servers for the replication of the communication between servers for the above-mentioned identification and a still more dynamic file is performed through the above-mentioned interface, an autonomous load distribution is realized effectively.
[0026] In addition, this invention is materialized also as invention concerning a method.
[0027]
[Embodiments of the Invention] Hereafter, with reference to a drawing, it explains per gestalt of operation of this invention.
[0028] [Operation gestalt of ** 1st] drawing 1 is the block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 1st operation gestalt of this invention.
[0029] In this drawing, it is the client (client computer) as which 1 requires a virtual distribution file server system of this file server system 1, and 2 requires file service of it. The virtual distribution file server system 1 is realized using the plurality (two sets of for example, file servers) (server computer) 10-1 distributed on the network 3, and 10-2. In addition, drawing, although only one set is shown for convenience, as for a client 2, it is common that more than one exist.
[0030] 11 is a virtual distributed file system which makes the center of the virtual distribution file server system 1, and is distributed and mounted in each file server 10-1 and 10-2. this virtual distributed file system 11 -- all the file servers 10-1 and the file of 10-2 -- integration ---like -- managing -- a file server 10-1 and 10-2 -- each actual volume composition (storage composition) is provided with the imagination file system for which it does not depend to a client 2
[0031] The virtual distributed file system 11 has each file server 10-1 and the virtual distribution file module 110-1,110-2 distributed and mounted in 10-2. The virtual distribution file module 110-1,110-2 is a management module for showing as one file system virtually to a client 2, distributing and processing the demand from a client 2 on a file server 10-1 and 10-2. The virtual distribution file module 110-1,110-2 The virtual distribution file interface 111-1,111-2 which processes nothing and the demand from a client 2 for the center of the module 110-1,110-2 concerned. The local file system 13-1 mentioned later and the interface 112-1,112-2 of 13-2 (local file interface). It has the communication module 113-1,113-2 which performs communication (communication between servers represented by the communication for the identification of the mapping table 14-1 mentioned later, the information on 14-2, and server

information) between the module 110-1,110-2 concerned.
[0032] A file server 10-1 and 10-2 are connected with the client 2 through the network 3. A file server 10-1 and 10-2 mount the storage equipments 12-1, such as a disk unit connected to the server 10-1 concerned and 10-2 other than the virtual distributed file system 11, the local file system (local file system) 13-1 which manages 12-2 (actual storage composition), and 13-2, respectively.
[0033] The virtual distributed file system 11, the local file system 13-1, the mapping table 14-1 of the same content which matches 13-2, and 14-2 are prepared in a file server 10-1 and 10-2. The data structure of this table 14-i (1 i= 2) is shown in drawing 2 .
[0034] The registration field 141 of the file name of a file where the virtual distributed file system 1 manages each entry of table 14-i (file name field), The logical whereabouts on the virtual distributed file system 1 of the file concerned Express. A path (It is visible from a client 2) The registration field of a (virtual path) (Virtual path field) 142, the registration field (whereabouts positional information field) 143 of the whereabouts positional information showing the physical whereabouts position on the storage of the file concerned (it is not visible from a client 2), and the access privilege (permission/prohibition) to the file concerned It has the registration field (permission information field) 144 of the permission information for managing, and the registration field 145 of other various attributes.
[0035] the virtual distributed file system 11 (upper virtual distribution file module 110-i) refers to mapping table 14-i of such a data structure -- for example, in a file server 10-1 and any of 10-2 a certain file's being and whereabouts information can be acquired, and also a permission etc. can acquire the attribute of a file if needed
[0036] The server information attaching part 15-1 of the same content and 15-2 are further prepared in a file server 10-1 and 10-2. Server information attaching part 15-i (1 i= 2) is used for holding server information including all the file servers 10-1 that constitute the virtual distribution file server system 1, the information (resource information) which shows the empty storage capacity (storage equipment 12-1 and 12-2) of 10-2, and the information which shows a load situation as shown in drawing 3 .
[0037] Next, operation of the composition of drawing 1 is explained. With this operation gestalt, it seems that each file server 10-1, the local file system 13-1 of 10-2, and not 13-2 but the virtual distributed file system 11 is mounted from the client 2. Then, a client 2 publishes the same demand to all the file servers 10-1 in which the virtual distributed file system 11 is mounted, and 10-2, when a certain file manipulation demand occurs. According to the technique of using IP (Internet Protocol) multicasting in this case, issue of a demand is possible for a client 2 side, without being conscious of the number of a file server.
[0038] A file server 10-1 and 10-2 will pass the demand concerned to the virtual distribution file module 110-1,110-2 corresponding to the self-server in the virtual distributed file system 11, if the demand from a client 2 is received. Then, a module 110-1,110-2 (inner virtual distribution file interface 111-1,111-2) is the read-out demand or the creation demand which it writes in (updating), and is a demand or is a creation demand or directory of a new file the demand of whose is a file, or distinguishes the demand classification.
[0039] Here, the file manipulation demand from a client 2 shall be a read-out demand or write request of a file. The file name of the file set as the object of a demand and the path on the virtual distributed file system 11 of the file concerned (virtual path) are given to this demand.
[0040] The virtual distribution file module 110-1,110-2 (inner virtual distribution file interface 111-1,111-2) When the file manipulation demand from a client 2 is a read-out demand or write request of a file, The file name and virtual path of a file which were demanded refer the mapping table 14-1 in a self-server, and 14-2. It investigates whether the file with the operation demand is held in the self-server (the storage equipment 12-1 connected to the self-server, 12-2) from the registration information on the table 14-1 with a file name and a virtual path concerned, and the whereabouts positional information field 143 in the entry in 14-2.
[0041] When the demanded file is held in the self-server, the virtual distribution file module 110-1,110-2 (inner virtual distribution file interface 111-1,111-2) accesses an actual file through the local file system 13-1 in a self-server, and 13-2 with the local file interface 112-1,112-2, and

returns a response to a client 2. On the other hand, when there is no file with the operation demand into a self-server, it is regarded as what other servers answer, and does not answer.
[0042] On the other hand, when the demand from a client 2 is creation of a new file, or creation of a directory, refer to the server (not being mapping table 14-1 and 14-2) information attaching part 15-1 and 15-2 for the virtual distribution file module 110-1,110-2. And according to a predetermined algorithm, only virtual distribution file module 110-i on [of any one] server 10-i (i is 1 or 2) receives the demand from a client 2 by virtual distribution file interface 111-i based on the server information on the server information attaching part 15-1 and all the servers currently held 15-2. Specifically, virtual distribution file module 110-i on server 10-i shall receive the demand from a client 2, when the empty storage capacity which the server information on all servers shows is measured and it can be judged with the empty storage capacity of self-server 10-i (storage equipment 12-i) being the largest. In this case, it is not necessary to necessarily give the information on a load situation into server information.
[0043] In addition, the load which the server information on all servers shows is compared, and you may make it the load of a self-server receive the demand from a client 2 to a low case most. In this case, it is not necessary to necessarily give the information on the empty storage capacity of a server [be / under / server information / correspondence / it].
[0044] In addition, it asks for each storage equipment 12-1 and a continuation field securable on 12-2 from the mapping information for every file of mapping table 14-i (from the operating condition of jamming each storage equipment 12-1 and the field of 12-2), and when storage equipment with the largest size is storage equipment 12-i of a self-server, you may make it for the continuation field more than required size to be securable, and receive the demand from a client 2. In this case, the server information attaching part 15-1 and 15-2 are not necessarily required.
[0045] Furthermore, it is good as for a method of judgment in whether it is the optimal server to which is asked for an evaluation value by making at least two of the sizes of empty storage capacity, a load situation, and a securable continuation field into conditions (compound condition), and a self-server receives a demand.
[0046] Now, in virtual distribution file module 110-i on file server 10-i, if the demand from a client 2 is received by virtual distribution file interface 111-i, local file system 13-i will perform creation of the demanded new file, or creation of a directory through local file interface 112-i, and the mapping table 14-1 and the entry information applicable to 14-2 will be registered.
[0047] After creation of a new file or creation of a directory is completed, in virtual distribution file module 110-i on file server 10-i, the new entry information registered into mapping table 14-i on a self-server is sent to virtual distribution file module 110-j on all other server 10-j (j is 1 or 2, however j!=i) through a network 3 by communication-module 113-i. Virtual distribution file module 110-j (inner virtual distribution file interface 111-j) receives the entry information on mapping table 14-i sent from virtual distribution file module 110-i through communication-module 113-j. And module 110-j (inner interface 111-j) received, and also registers the entry information of mapping table 14-i on server 10-i into mapping table 14-j in a self-server. Thus, the virtual distribution file module 110-1,110-2 on file server 10-1,110-2 can attain identification of the mapping table 14-1 concerned and the content of 14-2 by exchanging mutually the mapping table 14-1 and the entry information (entry information updated further) for which 14-2 was newly registered.
[0048] Moreover, a file server 10-1 and the virtual distribution file module 110-1,110-2 on 10-2 While updating periodically the server information on a self-server (empty storage capacity and load situation) among the server information attaching part 15-1 on a self-server, and the server information on each server currently held 15-2 By sending the updated server information to all other servers (upper virtual distribution file module 110-2,110-1) periodically through a network 3 (communication module 113-1,113-2) Identification of each file server 10-1, the server information attaching part 15-1 of 10-2, and the content of 15-2 is attained. That is, the virtual distribution file module 110-1,110-2 attains identification by exchanging server information mutually periodically.
[0049] By the above operation, distribution / cooperation operation of a file server 10-1 and 10-

2 can be carried out autonomously, and an imagination file server can be offered, without a file server making a client 2 in fact conscious of two sets (two or more sets) of a certain thing.
[0050] In addition, although the example of the system of drawing 1 explained the case where the number of servers was two, even if a server is three or more sets, an imagination file server can be offered according to the same structure.
[0051] [Operation gestalt of ** 2nd] drawing 4 is the block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 2nd operation gestalt of this invention, and has given the same sign to the same portion as drawing 1 .
[0052] In drawing 4 , the virtual distributed file system 41 which makes the center of the virtual distribution file server system 4 is distributed and mounted in n sets of file servers 10-1 -, and 10-n. this virtual distributed file system 41 -- the virtual distributed file system 11 in drawing 1 - - the same -- the file of all the file servers 10-1 - 10-n -- integration ---like -- managing -- each file server 10-1 - 10-n -- each actual volume composition is provided with the imagination file system for which it does not depend to a client 2 The virtual distributed file system 41 has the virtual distribution file module 410-1 which processes the demand from a client 2 on each file server 10-1 - 10-n - 410-n. A module 410-1 - 410-n have the virtual distribution file interface 411-1 which is the same composition as the module 110-1,110-2 in drawing 1 - 411-n, the local file interface 412-1 - 412-n, and a communication module 413-1 - 413-n. However, the communication module 413-1 in this operation gestalt - 413-n are constituted so that it may communicate through the private channel 5 mentioned later unlike the communication module 113-1,113-2 in drawing 1 .
[0053] A file server 10-1 - 10-n are connected with the client 2 through the network 3. A file server 10-1 - 10-n mount the local file system (local file system) 13-1 to 13-2 which manages the storage equipment 12-1 to 12-2 connected to the server 10-1 concerned - 10-n other than the virtual distributed file system 11, respectively. The mapping table 14-1 - 14-n, and the server information attaching part 15-1 to 15-2 are formed in a file server 10-1 - 10-n.
[0054] n sets of the point that the number of the file server from which the feature of the virtual distribution file server system 4 of the composition of drawing 4 constitutes a system unlike the virtual distribution file server system 1 of the composition of drawing 1 is n sets, and its file servers 10-1 -, and 10-n are the points by which interconnection is carried out also with private channel 5 with an another network 3. This private channel 5 is not specified about a physical layer, although it is Ethernet or a fiber channel (Fibre Channnel). Moreover, you may be a loop and a switch although the bus type is assumed in the example of drawing 4 also about topology.
[0055] In the composition of drawing 4 , in order for a file server 10-1 - 10-n to perform distribution / cooperation operation (the virtual distribution file module 410-1 in the virtual distributed file system 41 - 410-n) It is necessary to make the contents of the mapping table 14-1 - 14-n, and the server information attaching part 15-1 - 15-n always coincidence-ize among each server 10-1 - 10-n so that it may be guessed from explanation of operation with the operation form of the above 1st. However, when the number of the file server which constitutes the virtual distribution file server system 4 increases, the traffic (communication between servers of a sake) of the formation of information coincidence increases, and the throughput on a network 3 is made to get worse in performing information coincidence-ization between server 10-1 - 10-n through a network 3 like the operation form of the above 1st.
[0056] Then, with this operation gestalt (2nd operation gestalt), the private channel 5 only for [between servers] information interchange is formed like the composition of drawing 4 among each file server 10-1 ~ 10-n. To the communication between servers performed by a communication module 413-1 - 413-n by the virtual distribution file module 410-1 in the virtual distributed file system 41 - 410-n That is, it is made to use this channel 5 for the communication between servers for carrying out identification of the content of the mapping table 14-1 - 14-n, and the server information attaching part 15-1 - 15-n.
[0057] Thus, with this operation gestalt, mitigation of the load of a network 3 can be aimed at by not being a network 3 and using the private channel 5 for the communication between servers for the identification of the content of the mapping table 14-1 - 14-n, and the server information attaching part 15-1 - 15-n.
[0058] The 1st and 2nd operation gestalt described beyond [the 3rd operation gestalt] showed the example of the virtual distribution file server structure of a system which carries out distribution / cooperation operation of two or more file servers. The composition of this 1st [the], drawing 1 referred to with the 2nd operation gestalt, and drawing 4 is a static example in the specific number of a server. However, about the number of a server, considering as the composition which can be changed is desirable.
[0059] Then, the number of a server which constitutes a virtual distribution file server system is explained with reference to a drawing about the 3rd operation gestalt of this invention made dynamically extensible. Drawing 5 is the block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 3rd operation gestalt of this invention, and has given the same sign to the same portion as drawing 4 .
[0060] First, as shown in drawing 5 (a), new file server 10- (n+1) shall be added to the virtual distribution file server system 4 4 shown in drawing 4 , i.e., the virtual distribution file server system which consists of n sets of file servers 10-1 -, and 10-n.
[0061] In this case, virtual distribution file module 410- (n+1) on the virtual distributed file system 41 currently distributed by added file server 10- (n+1) To the file server 10-1 which already constitutes the virtual distribution file server system 4 - 10-n, as a sign A1 shows drawing 5 (a) (For example, the private channel which is not illustrated is minded) By communication between servers, renewal of the entry information on the mapping table 14-1 - 14-n and the server information on the server information attaching part 15-1 - 15-n (the resource information and load situation of each server are included) is locked.
[0062] Module 410- (n+1) on server 10- (n+1) moreover added As a sign A2 shows drawing 5 (b), from the server 10-1 of the either other file servers 10-1 or - the 10-n, for example, a file server All the information on the mapping table 14-1 and the server information attaching part 15-1 is copied to mapping table 14- in a self-server (n+1), and server information attaching part 15- (n+1) by communication between servers.
[0063] Next, to server information attaching part 15- after a copy (n+1), module 410- (n+1) on added file server 10- (n+1) adds the server information which shows the resource and load situation of a self-server, as a sign A3 shows drawing 5 (c).
[0064] After an appropriate time, as a sign A4 shows drawing 5 (d), module 410- (n+1) on file server 10- (n+1) publishes an identification demand of server information to all other file servers 10-1 - 10-n by communication between servers after an appropriate time, and cancels a lock after that.
[0065] By a series of above operation, a firewood shelf server (file server 10- (n+1)) can be dynamically added to the virtual distribution file server system 4 already built. If the information how a new resource is distributed to the volume composition of the present virtual distribution file server system 4 in this case is added, it is also possible to extend volume alternatively if needed.
[0066] [Operation gestalt of ** 4th] drawing 6 is the block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 4th operation gestalt of this invention, and has given the same sign to the same portion as drawing 4 .
[0067] In drawing 6 , 6 is a virtual distribution file server system equivalent to the virtual distribution file server system 4 in drawing 4 . The feature of this virtual distribution file server system 6 is equipped with the load status information attaching part 16-1 classified by file holding the information (load status information classified by file) on the load situation about each file currently held at the self-server (storage equipment 12-1 - 12-n) - 16-n in the file server 10-1 which constitutes the system 6 concerned - 10-n. In connection with this, the functions which the virtual distributed file system 61 which makes the center of the virtual distribution file server system 6 has also differ in part in the virtual distributed file system 41 in drawing 4 . However, the same sign (410-1 - 410-n) as drawing 4 is used for the virtual distribution file module of each file server 10-1 in the virtual distributed file system 61 - every

10-n for convenience. In addition, in drawing 6 , the component in a module 410-1 - 410-n (a virtual distribution file interface, a local file interface, communication module) and the mapping table on a file server 10-1 - 10-n, and the server information attaching part are omitted.
[0068] Load status information attaching part 16classified by file-i (i=1-n) has the data structure shown in drawing 7 (a), and holds the load status information classified by file containing the information which shows the load situation for every file in a self-server, the information (file attribute) which shows the attribute of the file, and a replication flag. It is shown whether the corresponding file of a file attribute is original or it is a replica (duplicate). Moreover, a replication flag shows whether it is whether it being generation ending about the replica at another servers side, and replication ending that is, when a corresponding flag is original.
[0069] Signs that the replica 612 of the arbitrary files 611 in the storage equipment 12-1 which a file server 10-1 has in storage equipment 12-n which file server 10-n has is held by the replication B1 are shown in drawing 6 .
[0070] Next, operation of the composition of drawing 6 is explained. Refer to the load status information attaching part 16-1 classified by file - the 16-n for each virtual distribution file module 410-1 on the virtual distributed file system 61 - 410-n periodically, for example. And a module 410-1 - 410-n From the load status information according to file currently held at an attaching part 16-1 - 16-n In the file currently held at the self-server (storage equipment 12-1 - 12-n) When it is detected that the file of the load beyond the 1st threshold exists, communication between servers through the private channel 5 performs replication operation which generates asynchronously the replica of a file which corresponds to one of the other servers. The load situation of a file is total of the number of demands in the queue (queue) of the demand to the file concerned, or the size which the demand in the waiting state of the file concerned shows, and whenever it finishes processing a demand as whenever it receives a demand, it is updated here. Moreover, based on the server information currently held at the server information attaching part which is not illustrated, a load should just choose a low server as the object server of a replication most.
[0071] The virtual distribution file module 410-1 - 410-n will set to a notice state [finishing / a replication] the replication flag in the load status information of the corresponding file currently held at the load status information attaching part 16-1 classified by file of a self-server - 16-n, if replication operation is performed. Moreover, the virtual distribution file module of a server set as the object of replication operation adds the load status information of a replica [in the load status information attaching part classified by file of a self-server].
[0072] Here, as shown in drawing 6 , the replication B1 of the file 611 which a file server 10-1 holds should be performed to file server 10-n through the private channel 5, and the replica 612 should be held at storage equipment 12-[of the file server 10-n concerned] n. In this case, the replication flag in the load status information of the file 611 currently held at the load status information attaching part 16-1 classified by file of a file server 10-1 is set to the state which shows replication ending. Moreover, the new load status information about the replica 612 of a file 611 is added to load status information attaching part 16classified by file-[of file server 10-n] n. The file attribute in this load status information shows that a corresponding file is a replica (file 611) (612).
[0073] Henceforth, when there is a new read-out demand of a file 611 from a client 2, if the file server 10-1 (virtual distribution file module 410-1) holding the file 611 concerned investigated whether the load of the file 611 concerned would be over the 2nd threshold (however, the 2nd threshold < 1st threshold) and has exceeded, it will not answer a demand from a client 2. In this case, to the demand from a client 2, file server 10-n which received the replication answers. File server 10-n does not need to take into consideration whether a file server 10-1 answers, and as long as it has the replica 612 of the demanded file 611, it should just answer a client 2 here.
[0074] Thus, by file server 10-n processing the new read-out demand to the file 611 from a client 2 using the replica 612, by the file server 10-1 holding the file 611, processing of the read-out demand to the file 611 concerned received before it progresses, and the load of the file 611 concerned becomes below the threshold of the above 2nd. Then, the virtual distribution file module 410-1 on a file server 10-1 sends the demand for eliminating the replica 612 of a file 611

by communication between servers through the private channel 5 to virtual distribution file module 410-n on file server 10-n.
[0075] Virtual distribution file module 410-n on file server 10-n which received this demand processes using a replica 612 only to a received demand already, and eliminates the load status information corresponding to after an appropriate time with a replica 612. On the other hand, if the virtual distribution file module 410-1 on a file server 10-1 has the new read-out demand to the file 611 from a client 2, it will answer to it.
[0076] By the way, as a result of coming to receive the read-out demand to the file 611 concerned by file server 10-n, before the load of the file 611 in a file server 10-1 becomes below the 2nd threshold by the replication of the file 611 from a file server 10-1 to file server 10-n, the load of the replica 612 of the file 611 concerned in file server 10-n can exceed the 1st threshold.
[0077] Then, what is necessary is for file server 10-n to generate the next generation's replica to other one server using a replica 612 shortly, namely, to perform the replication of a replication, and just to make the read-out demand to a file 611 process by the server in such a case. for that purpose, a load situation and a replication flag as shown in the load status information for every file held at load status information attaching part 16classified by file-i (i=1-n) at drawing 7 (a) -- in addition, as shown in drawing 7 (b), it is good to give the generation information on a file
[0078] In this case, when the load of a certain generation's replica falls below in the threshold of the above 2nd, it is controllable to eliminate the replica of the next generation which the server set as the object of the replication by the server has from a server with the replica concerned etc. the case where the server as which elimination of a replica was required is generating the replica of the next generation further to another server at this time -- the -- it is good to eliminate the replica of the next generation further In addition, you may make it the server in which a load has a low file most answer the read-out demand to the file concerned about the same file (for a replica to be included) by attaining identification between each server at least like the server information described above about the load status information of the file relevant to the replica.
[0079] recently -- streaming data, such as video and an audio, -- or the contents of WWW (World Wide Web) etc. have fundamentally main read-out, its size is comparatively large, and the data which need a certain amount of response (it is a band guarantee depending on the case) are increasing them And since the case which looks at in the short term and access concentrates on specific data (file) is assumed, such data may be difficult to secure a response. The composition of drawing 6 described above is a thing supposing such a situation, and when access concentrates on a specific file, it enables it to distribute access of the FAIRUHE concerned by performing REBURIKESHON of the file concerned automatically. Not only a load distribution but this composition can be used for backup of the high file of importance.
[0080] [Operation gestalt of ** 5th] drawing 8 is the block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 5th operation gestalt of this invention, and has given the same sign to the same portion as drawing 4 .
[0081] In drawing 8 , 8 is a virtual distribution file server system equivalent to the virtual distribution file server system 4 in drawing 4 . The interconnection of a file server 10-1 - 10-n, and storage equipment 12-1 - 12-n is carried out by FC-AL (Fibre Channnel Arbitrated Loop)80, and the feature of this virtual distribution file server system 8 is that it has applied the network configuration (that is, a multi-host is possible) which can share storage equipment (as a target) 12-1 - 12-n from each file server (as a host) 10-1 - 10-n. Here, unlike the composition of drawing 4 , be careful of a point without the private channel 5.
[0082] What is necessary is just to perform communication between servers performed through the private channel 5 in the composition of drawing 4 (the communication module 413-1 of the virtual distribution file module 410-1 - 410-n - 413-n) through a network 3 with the composition of this drawing 8 like the composition of drawing 1 (as for drawing, this state is shown). Moreover, you may be made to perform the above-mentioned communication between servers

on FC-AL80 through the interface for storage connection of a file server 10-1 - 10-n. In this case, the load of a network 3 is mitigable the same with having used the private channel 5.
[0083] According to the composition of drawing 8 , since storage equipment 12-1 - 12-n can be directly seen from all the file servers 10-1 - 10-n, replication operation and a load distribution which were stated with the operation gestalt of the above 4th can be easily performed by giving the load status information attaching part 16-1 classified by file in drawing 6 - 16-n to each server 10-1 - 10-n. In addition, the network (interface) in which a multi-host is possible may not be restricted to FC-AL80, and may be a SCSI (Small Computer System Interface) bus.
[0084]
[Effect of the Invention] As explained in full detail above, according to this invention, from a client, he can treat two or more file servers distributed on the network as a single server, and a client is not made conscious of the connection state of the number of a server, or storage equipment.
[0085] Moreover, according to this invention, when a server is extended, volume can also be extended dynamically.
[0086] Furthermore, according to this invention, an autonomous load distribution is realizable among two or more servers.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] The block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 1st operation gestalt of this invention.
[Drawing 2] Drawing showing the example of a data structure of the mapping table in drawing 1 .
[Drawing 3] Drawing showing the example of a data structure of the server information attaching part in drawing 1 .
[Drawing 4] The block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 2nd operation gestalt of this invention.
[Drawing 5] The block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 3rd operation gestalt of this invention.
[Drawing 6] The block diagram showing the composition of the computer network system which applies the virtual distribution file server system concerning the 4th operation gestalt of this invention.
[Drawing 7] Drawing showing the example of a data structure of the load status information attaching part classified by file in drawing 6 .
[Drawing 8] The timing chart explaining operation of this operation gestalt.
[Description of Notations]
1, 4, 6, 8 -- Virtual distribution file server system
2 -- Client
3 -- Network
5 -- Private channel
10-1 - 10-n -- File server
11, 41, 61 -- Virtual distributed file system
12-1 - 12-n -- Storage equipment
13-1 - 13-n -- Local file system
14-1 - 14-n -- Mapping table
15-1 - 15-n -- Server information attaching part
16-1 - 6-n -- Load status information attaching part classified by file
80 -- FC-AL (interface in which a multi-host is possible)
110-1 - 110-n, 410-1 - 410-n -- Virtual distribution file module (management module)
111-1 - 111-n -- Virtual distribution file interface
112-1 - 112-n -- Local file interface
113-1 - 113-n -- Communication module
611 -- File
612 -- (file 611) Replica

[Translation done.]